COMP90042

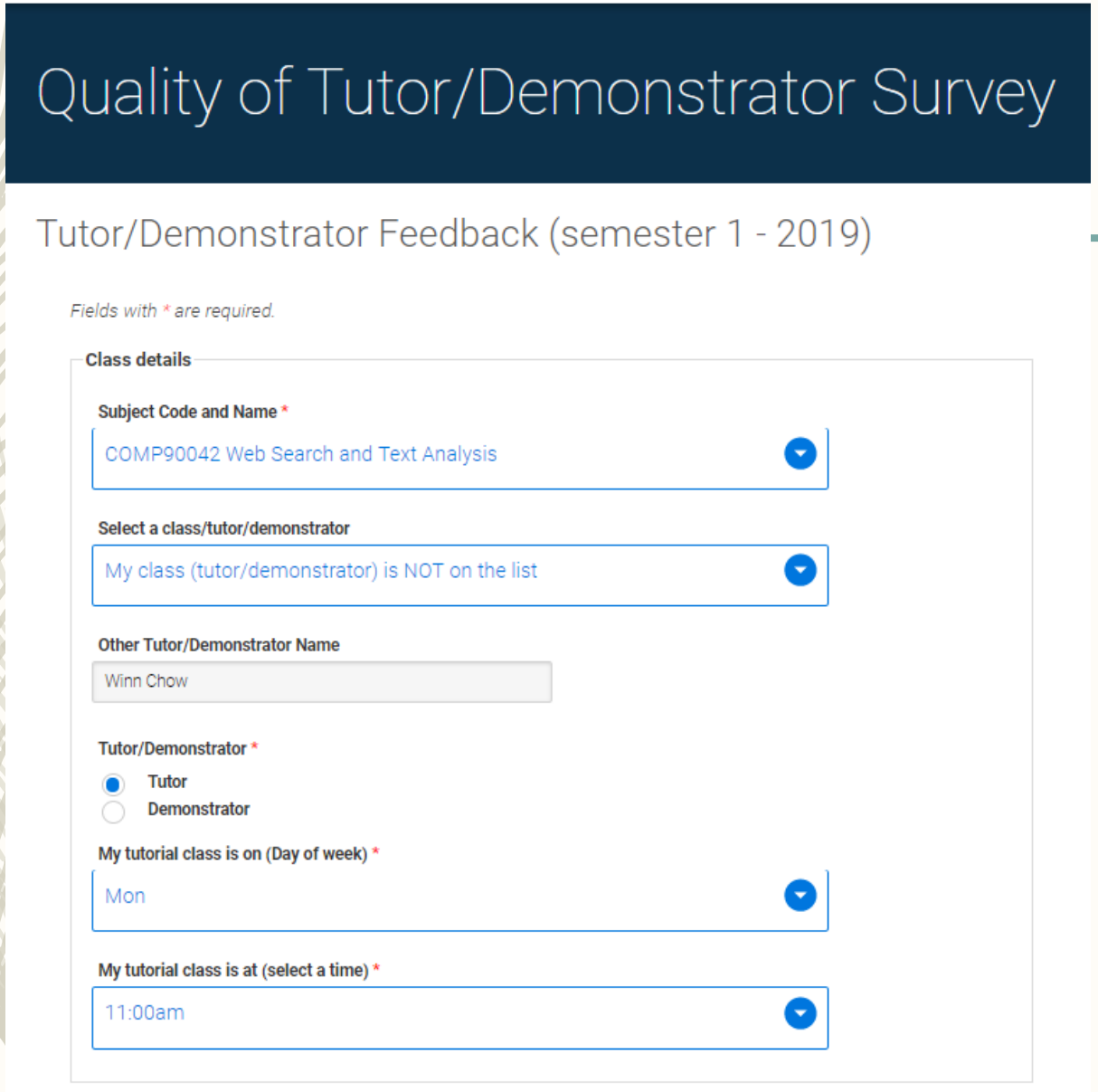
Web search and text analysis

Workshop Week 12

Your tutor

- Winn Chow (Senior Tutor)
- winn.chow1@unimelb.edu.au
- Office: Doug McDonell - 9.23
- Here, you can find my workshop slides:
- https://github.com/winnchow/COMP90042-Workshops

https://apps.eng.unimelb.edu.au/casmas/index.php?r=qoct/subjects



My class is NOT on the list

Tutor: Winn Chow

Mon 11:00am

Q1 and Q2

1. What aspects of human language make automatic translation difficult?
2. For the following "bi-text":

Language A	Language B
`green house`	`casa verde`
`the house`	`la casa`

(a) What is the logic behind **IBM Model 1** for deriving word alignments?

(b) Work through the first 2 iterations of the **Expectation Maximisation** algorithm to build a translation table for this collection, based on IBM Model 1. Check your work by comparing to the `WSTA_N21_machine_translation.ipynb` output.

Translation is hard

However , the sky remained clear under the strong north wind .

虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。

Although north wind howls , but sky still extremely limpid .

- Not just simple word for word translation
  - structural changes, e.g., syntax and semantics
  - multiple word translations, idioms
  - inflections for gender, case etc
  - missing information (e.g., determiners)

Example from Lopez, 2008, PhD dissertation UMD

Translating $f \rightarrow e$, P(e | f)

Noisy channel MT

- Two components:

Translation Model (TM)

$$\hat{e} = \text{argmax}_e \, P(e)P(f|e)$$

Language Model (LM)

- Responsible for:
  - *P(f|e)* rewards good translations, but permissive of disfluent *e*
  - *P(e)* rewards *e* which look like fluent English, and helps put words in the correct order

Q: Why not just one TM to model *P(e|f)* directly?

Translating $f \rightarrow e$, P(e | f)

- Need to calculate expected alignments under the model (step 2)

$$P(\mathbf{a}|\mathbf{e},\mathbf{f}) = \frac{P(\mathbf{f},\mathbf{a},\mathbf{e})}{P(\mathbf{f},\mathbf{e})} = \frac{P(\mathbf{f},\mathbf{a}|\mathbf{e})}{P(\mathbf{f}|\mathbf{e})}$$

Representing Alignment

- Representation:

  $E = e_1 \ldots e_I$ = *And the program has been implemented*

  $F = f_1 \ldots f_J$ = *Le programme a ete mis en application*

  $A = a_1 \ldots a_J$ = 2, 3, 4, 5, 6, 6, 6.

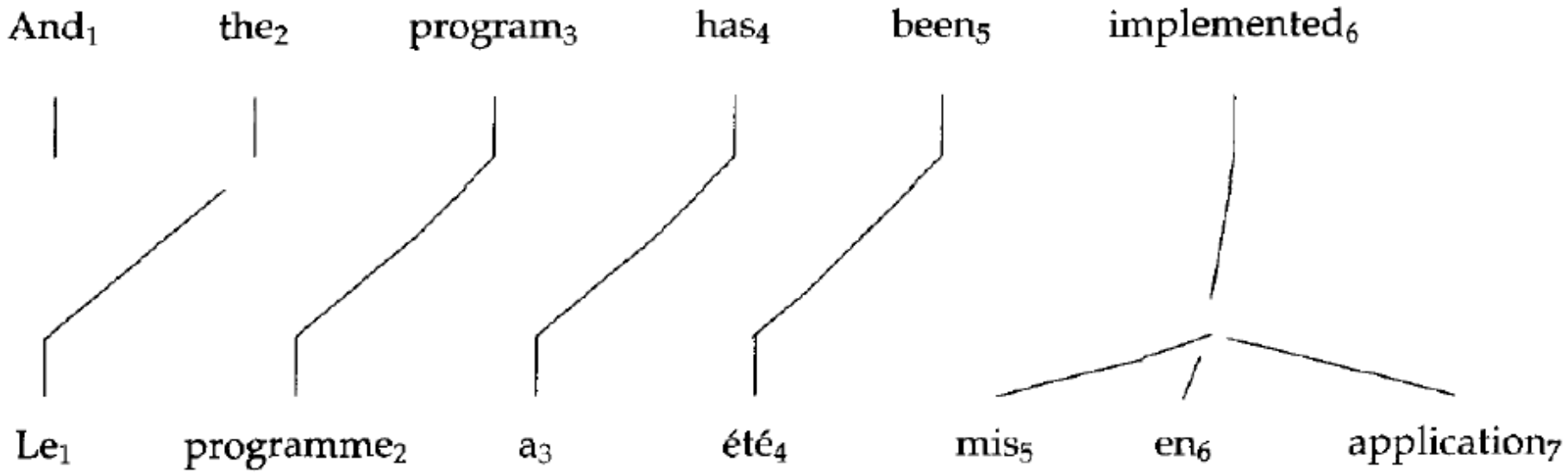


Figure from Brown, Della Pietra, Della Pietra, Mercer, 1993

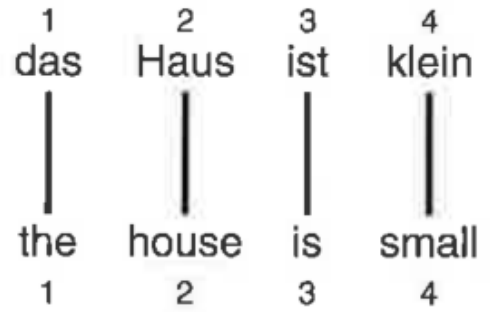
Example IBM

- Given translation table, evaluate the probability of the aligned sentence pair

e = the		e = house		e = is		e = small	
f	**t(f|e)**	**f**	**t(f|e)**	**f**	**t(f|e)**	**f**	**t(f|e)**
das	0.4	Haus	0.35	ist	0.2	klein	0.4
der	0.35	Geschlect	0.05	bin	0.15	gering	0.25
die	0.25	Häuser	0.15	bist	0.10	schmal	0.15
		aufnehmen	0.20	sein	0.30		
		Heim	0.25	sind	0.25		

$$P(\mathbf{f}, \mathbf{a}|\mathbf{e}) = \frac{\epsilon}{5^4} t(\text{das}|\text{the}) t(\text{Haus}|\text{house}) t(\text{ist}|\text{is}) t(\text{klein}|\text{small})$$
$$= 0.00029\epsilon$$

Example adapted from Koehn 09



Translate B → A

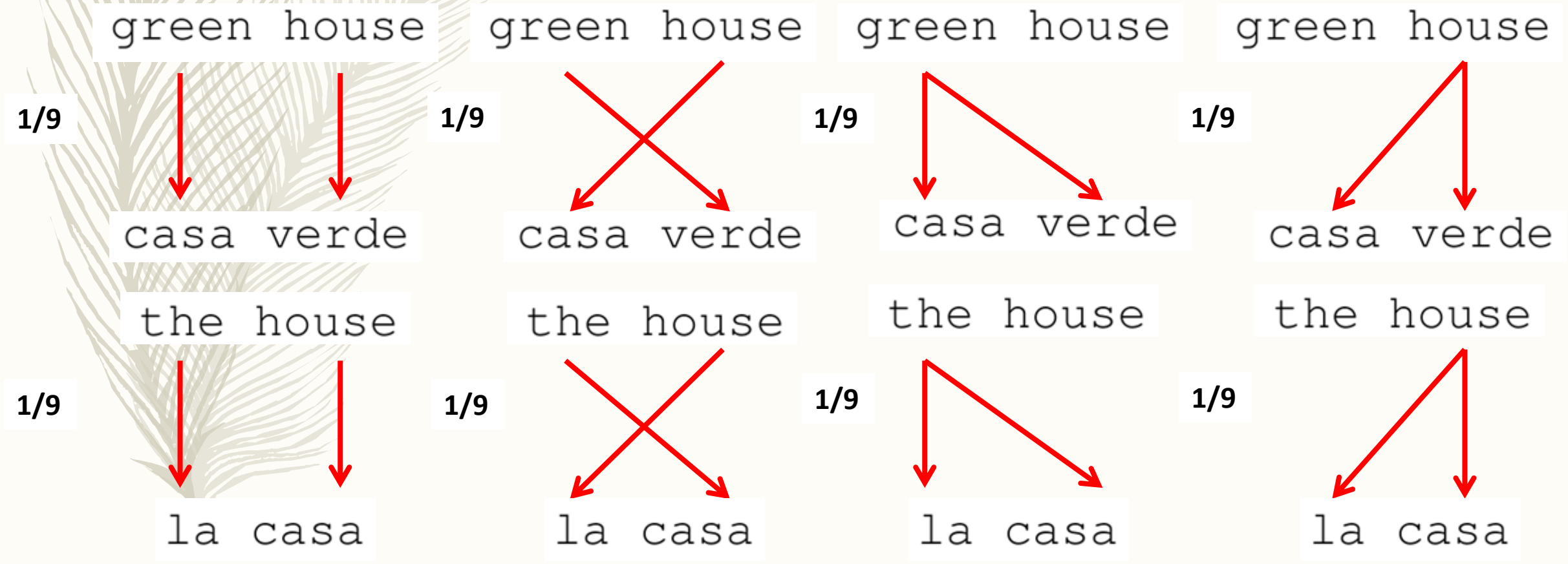
Language A	Language B
green house	casa verde
the house	la casa

1. make initial guess of *t* parameters, e.g., uniform

t(B|A)	casa	la	verde	Total
green	1/3	1/3	1/3	1
house	1/3	1/3	1/3	1
the	1/3	1/3	1/3	1

- Need to calculate expected alignments under the model (step 2)

$$P(\mathbf{a}|\mathbf{e},\mathbf{f}) = \frac{P(\mathbf{f},\mathbf{a},\mathbf{e})}{P(\mathbf{f},\mathbf{e})} = \frac{P(\mathbf{f},\mathbf{a}|\mathbf{e})}{P(\mathbf{f}|\mathbf{e})}$$



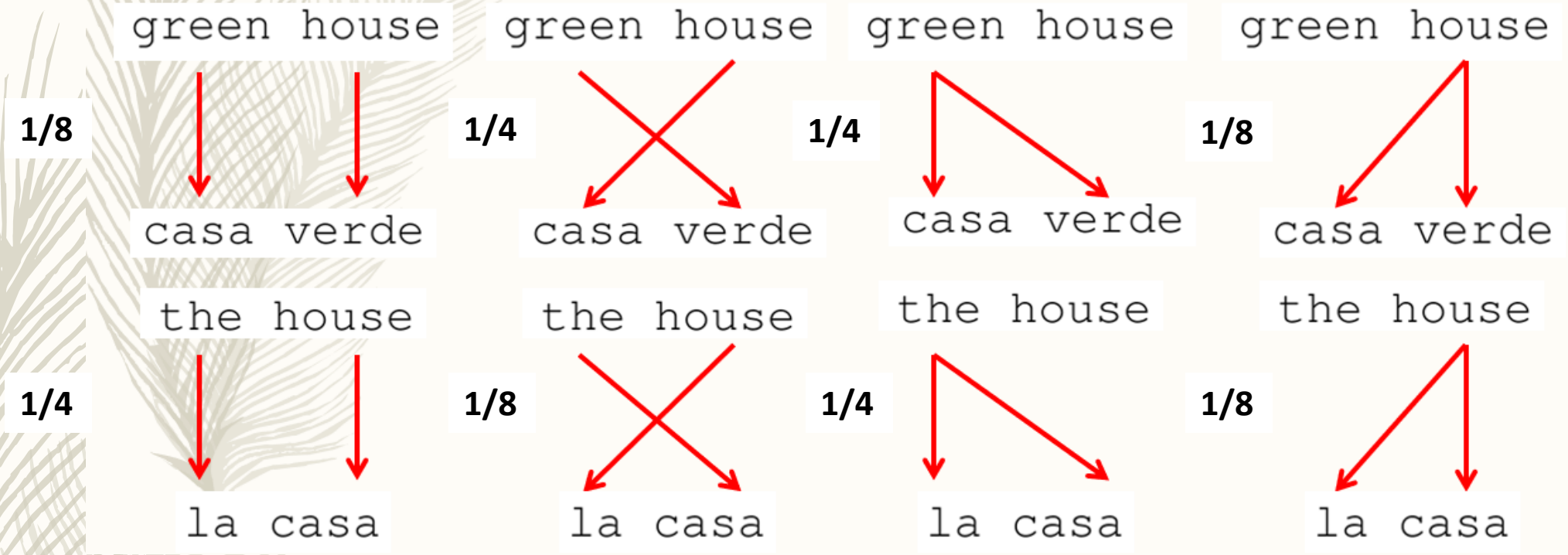
- Need to calculate expected alignments under the model (step 2)

$$P(\mathbf{a}|\mathbf{e},\mathbf{f}) = \frac{P(\mathbf{f},\mathbf{a},\mathbf{e})}{P(\mathbf{f},\mathbf{e})} = \frac{P(\mathbf{f},\mathbf{a}|\mathbf{e})}{P(\mathbf{f}|\mathbf{e})}$$

$$\begin{aligned}\hat{P}(F,A|E) &= \frac{\epsilon}{(I+1)^J} t(\texttt{casa}|\texttt{green}) t(\texttt{verde}|\texttt{house}) \\ &= \frac{\epsilon}{(2+1)^2}\left(\frac{1}{3}\right)\left(\frac{1}{3}\right) = \frac{\epsilon}{9}\frac{1}{9}\end{aligned}$$

(ignoring the ϵ term):

t(B|A)	casa	la	verde	Total
green	1/9 * 2	0	1/9 * 2	4/9
house	1/9 * 4	1/9 * 2	1/9 * 2	8/9
the	1/9 * 2	1/9 * 2	0	4/9



t(B|A)	casa	la	verde	Total
green	1/2	0	1/2	1
house	1/2	1/4	1/4	1
the	1/2	1/2	0	1

- For Ia, we observe the following (ignoring the ϵ term):

$$\begin{aligned}\hat{P}(F, A|E) &= t(\texttt{casa}|\texttt{green})t(\texttt{verde}|\texttt{house}) \\ &= (\frac{1}{2})(\frac{1}{4}) = \frac{1}{8}\end{aligned}$$

- For Ib:

$$\begin{aligned}\hat{P}(F, A|E) &= t(\texttt{verde}|\texttt{green})t(\texttt{casa}|\texttt{house}) \\ &= (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}\end{aligned}$$

- For Ic:

$$\begin{aligned}\hat{P}(F, A|E) &= t(\texttt{casa}|\texttt{green})t(\texttt{verde}|\texttt{green}) \\ &= (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}\end{aligned}$$

- For Id:

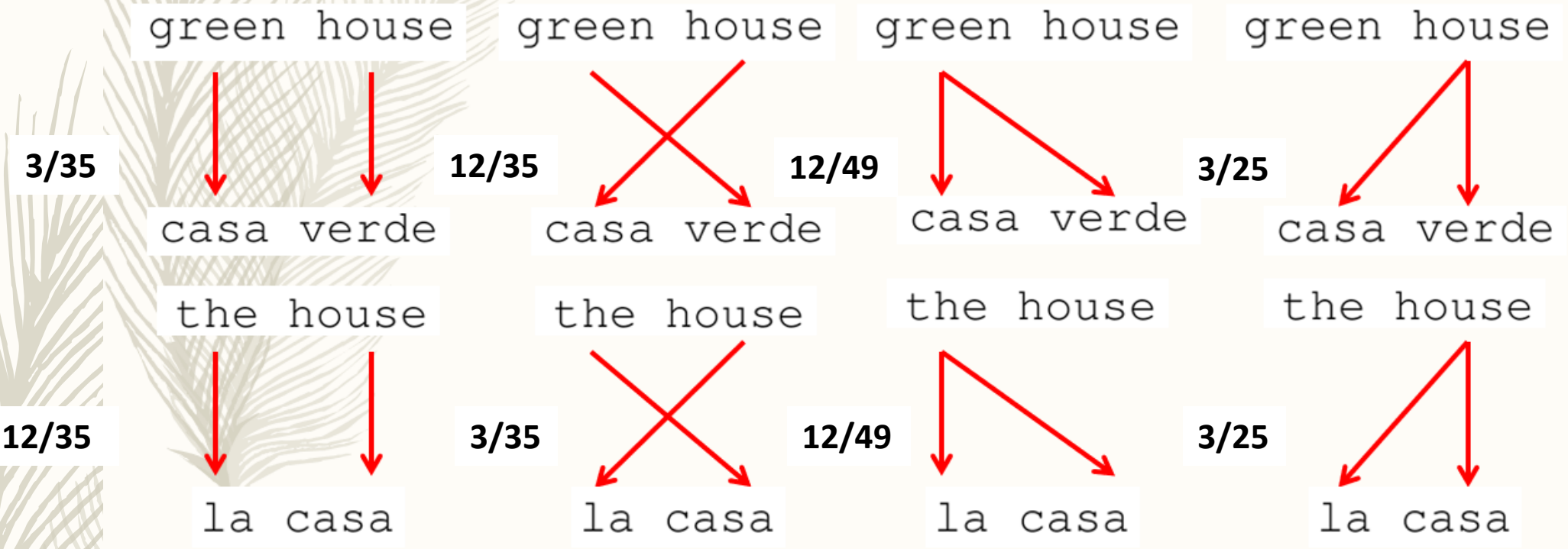
$$\begin{aligned}\hat{P}(F, A|E) &= t(\texttt{casa}|\texttt{house})t(\texttt{verde}|\texttt{house}) \\ &= (\frac{1}{2})(\frac{1}{4}) = \frac{1}{8}\end{aligned}$$

- Need to calculate expected alignments under the model (step 2)

$$P(\mathbf{a}|\mathbf{e},\mathbf{f}) = \frac{P(\mathbf{f},\mathbf{a},\mathbf{e})}{P(\mathbf{f},\mathbf{e})} = \frac{P(\mathbf{f},\mathbf{a}|\mathbf{e})}{P(\mathbf{f}|\mathbf{e})}$$

(ignoring the ϵ term):

t(B|A)	casa	la	verde	Total
green	1/8 + 1/4	0	1/4 + 1/4	7/8
house	1/4 + 1/8 + 1/4 + 1/8	1/8 + 1/8	1/8 + 1/8	10/8
the	1/8 + 1/4	1/4 + 1/4	0	7/8



t(B|A)	casa	la	verde	Total
green	3/7	0	4/7	1
house	3/5	1/5	1/5	1
the	3/7	4/7	0	1

- Further iterations will continue to improve these counts, and to observe that Ib and IIa are the most likely alignments.